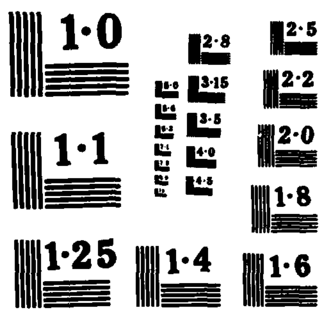
AD-A145 810 CONVERGENCE OF SIP (STRONGLY IMPLICIT PROCEDURE)(U) 1/1
YALE UNIV NEW HAVEN CT DEPT OF COMPUTER SCIENCE P CHEN
MAY 84 YALEU/DCS/RR-321 N00014-82-K-0184

UNCLASSIFIED F/G 12/1 NL

END
DATE
FILMED
10 84


1·0
2·8
2·5
3·15
2·2
3·5
1·1
4·0
2·0
4·5
1·8
1·25
1·4
1·6

AD-A145 810

12

DTIC FILE COPY

YALE UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

84 07 24 020

ABSTRACT

We prove that the Strongly Implicit Procedure ~~(SIP), due to H. L. Stone,~~ for solving self-adjoint elliptic difference equations is convergent.

Convergence of SIP

Peixian Chen †

Research Report YALEU/DCS/RR-321
May 1984

— A —

†Department of Computer Science, Yale Univ. and Department of Mathematics, Beijing Univ., Beijing, China.

This work was supported in part by The U.S. Office of Naval Research under ONR Grant N00014-82-K-0184.

1. Introduction

Consider the following elliptic boundary problem

$$\begin{cases} -\dfrac{\partial}{\partial x}p_1(x,y)\dfrac{\partial u}{\partial x} - \dfrac{\partial}{\partial y}p_2(x,y)\dfrac{\partial u}{\partial y} = g(x,y), & (x,y) \in D = \{(x,y) \mid 0 < x,y, < 1\}; \\ u(x,y) = r(x,y), & (x,y) \in \partial D; \end{cases} \tag{1.1}$$

where $p_i(x,y) \geq p_0 > 0$, $i = 1,2$ and p_0 is a constant. Approximating (1.1) by a five-point finite-difference scheme on the grid system $D_h = \{(jh,kh) | \quad h = 1/(n+1); \quad j,k = 1,2,\ldots,n\}$, we obtain a set of liear algebric equations

$$Au = Q, \tag{1.2}$$

where u and Q are n^2-column vectors whose components are grid functions $u_{j,k}$ and $q_{j,k}$, respectively, ordered in a left-to-right and down-to-up fashion, and the matrix A is defined by the following *

$$(Au)_{j,k} = B_{j,k}u_{j,k-1} + D_{j,k}u_{j-1,k} + E_{j,k}u_{j,k} + F_{j,k}u_{j+1,k} + G_{j,k}u_{j,k+1}, \qquad 1 \leq j,k \leq n, \tag{1.3}$$

whose coefficients satisfy the conditions

$$\begin{cases} (1) B_{j,k}, D_{j,k}, F_{j,k} \text{ and } G_{j,k} \text{ are all less than a negative constant if} \\ \qquad \text{subscripts of their associated grid functions are in [1,n];} \\ (2) E_{j,k} + B_{j,k} + D_{j,k} + F_{j,k} + G_{j,k} \begin{cases} > 0, & \text{if } B_{j,k} \cdot D_{j,k} \cdot F_{j,k} \cdot G_{j,k} = 0 \\ = 0, & \text{otherwise;} \end{cases} \end{cases} \tag{1.4}$$

and

$$D_{j+1,k} = F_{j,k}, \qquad B_{j,k+1} = G_{j,k}. \tag{1.5}$$

It is well known that A is a symmetric positive-definite matrix with five nonzero diagonals.

In [4], H. L. Stone proposed an iterative method, Strongly Implicit Procedure (SIP), for solving (1.2). First, he approximately factored A as a product of a lower triangular matrix L and an upper triangular matrix U which have the same sparsity as the lower and the upper triangular parts of A, respectively, i.e.

$$\begin{aligned} (Lu)_{j,k} &= b_{j,k}u_{j,k-1} + c_{j,k}u_{j-1,k} + d_{j,k}u_{j,k}, \\ (Uu)_{j,k} &= u_{j,k} + e_{j,k}u_{j+1,k} + f_{j,k}u_{j,k+1}, \end{aligned} \qquad 1 \leq j,k \leq n, \tag{1.6}$$

so that

$$\begin{aligned} ((LU)u)_{j,k} = {} & b_{j,k}u_{j,k-1} + b_{j,k}e_{j,k-1}u_{j+1,k-1} + c_{j,k}u_{j-1,k} \\ & + (d_{j,k} + b_{j,k}f_{j,k-1} + c_{j,k}e_{j-1,k})u_{j,k} \\ & + d_{j,k}e_{j,k}u_{j+1,k} + c_{j,k}f_{j-1,k}u_{j-1,k+1} + d_{j,k}f_{j,k}u_{j,k+1}. \end{aligned}$$

Let the matrix B be defined as follows:

$$\begin{aligned} (Bu)_{j,k} = {} & b_{j,k}e_{j,k-1}\left\{u_{j+1,k-1} + \alpha u_{j,k} - \alpha u_{j+1,k} - \alpha u_{j,k-1}\right\} \\ & + c_{j,k}f_{j-1,k}\left\{u_{j-1,k+1} + \alpha u_{j,k} - \alpha u_{j-1,k} - \alpha u_{j,k+1}\right\}. \end{aligned} \qquad 1 \leq j,k \leq n \tag{1.7}$$

*Throughout this paper, as long as a matrix is defined in this way we make a convention that if the subscripts of a grid function are not in [1,n] then the terms which include such grid function do not exist, or coefficients are zero.

The nonzero elements of L and U are defined so that $A + B = LU$. The algorithm for computing these elements is

$$\begin{aligned} b_{j,k} &= B_{j,k} - \alpha b_{j,k} e_{j,k-1}, \\ c_{j,k} &= D_{j,k} - \alpha c_{j,k} f_{j-1,k}, \\ d_{j,k} + b_{j,k} f_{j,k-1} + c_{j,k} e_{j-1,k} &= E_{j,k} + \alpha b_{j,k} e_{j,k-1} + \alpha c_{j,k} f_{j-1,k}, \qquad 1 \le j,k \le n, \alpha \in [0,1] \\ d_{j,k} e_{j,k} &= F_{j,k} - \alpha b_{j,k} e_{j,k-1}, \\ d_{j,k} f_{j,k} &= G_{j,k} - \alpha c_{j,k} f_{j-1,k}, \end{aligned} \tag{1.8}$$

The SIP iterative scheme is given by

$$(A + B)(u^{(s+1)} - u^{(s)}) = \tau(Q - Au^{(s)})\dagger \qquad s = 0, 1, \cdots. \tag{1.9}$$

The practical numerical computations given in [4] suggested that SIP is convergent and more efficient than the standard iterative methods. After Stone the theoretical analyses of the approximate factorization iterative methods given by T. Dupont, R. Kendall, H. Rachford[2], A. Bracha-Barak and P. Saylor[1] etc. were important contributions to the later development of the more efficient preconditioned conjugate gradient. However, these studies did not prove that SIP is convergent and SIP was layed aside. The purpose of this paper is to analyze the convergence of SIP by the methods of those mentioned above. In Section 2 we discuss a property of $A + B$, where A is not confined to be a symmetric matrix. In Section 3 the convergence of SIP is proved.

2. The property of $A + B$

As in Lemma 1 in [1] we have the following theorem.

Theorem 2.1. *If the matrix A satisfies (1.4) and $0 \le \alpha \le 1$ then the quantities $b_{j,k}$, $c_{j,k}$, $d_{j,k}$, $e_{j,k}$, and $f_{j,k}$ defined by (1.8) satisfy*

$$\begin{aligned} &(1)\ b_{j,k} \le B_{j,k} \quad \text{and} \quad c_{j,k} \le D_{j,k}, \\ &(2)\ d_{j,k} > 0, \\ &(3)\ -1 < e_{j,k} \le F_{j,k}/d_{j,k} \quad \text{and} \quad -1 < f_{j,k} \le G_{j,k}/d_{j,k}, \\ &(4)\ 1 + e_{j,k} + f_{j,k} > 0, \end{aligned} \tag{2.1}$$

and they are all continuous functions of $\alpha \in [0,1]$.

Proof. The proof is by induction. Obviously, the theorem holds for $j = k = 1$. Now assume the conclusions of the theorem hold for all points preceding (j, k). First, we have

$$b_{j,k} = B_{j,k}/(1 + \alpha e_{j,k-1}) \le B_{j,k},$$

$$c_{j,k} = D_{j,k}/(1 + \alpha f_{j-1,k}) \le D_{j,k},$$

and

$$\begin{aligned} d_{j,k} &= E_{j,k} - b_{j,k} f_{j,k-1} - c_{j,k} e_{j-1,k} + \alpha b_{j,k} e_{j,k-1} + \alpha c_{j,k} f_{j-1,k} \\ &= E_{j,k} - b_{j,k} f_{j,k-1} - c_{j,k} e_{j-1,k} + (B_{j,k} - b_{j,k}) + (D_{j,k} - c_{j,k}) \\ &= E_{j,k} + B_{j,k} + D_{j,k} - b_{j,k}(1 + f_{j,k-1}) - c_{j,k}(1 + e_{j-1,k}) > 0. \end{aligned} \tag{2.2}$$

† In [4], $\tau = 1$.

Since

$$\begin{aligned} d_{j,k} + d_{j,k}e_{j,k} &= E_{j,k} + B_{j,k} + D_{j,k} - b_{j,k}(1 + f_{j,k-1}) \\ &\qquad - c_{j,k}(1 + e_{j-1,k}) + F_{j,k} - \alpha b_{j,k}e_{j,k-1} \\ &= E_{j,k} + B_{j,k} + D_{j,k} + F_{j,k} \\ &\qquad - b_{j,k}(1 + \alpha e_{j,k-1} + f_{j,k-1}) \\ &\qquad - c_{j,k}(1 + e_{j-1,k}) > 0 \end{aligned}$$

then

$$1 + e_{j,k} > 0,$$

i.e.

$$-1 < e_{j,k} = (F_{j,k} - \alpha b_{j,k}e_{j,k-1})/d_{j,k} \leq F_{j,k}/d_{j,k}.$$

Similarly,

$$-1 < f_{j,k} \leq G_{j,k}/d_{j,k}.$$

In the above expressions, both numerators and denominators are continuous functions of $\alpha \in [0,1]$ and the denominators are nonzero. So $b_{j,k}$, $c_{j,k}$, $d_{j,k}$, $e_{j,k}$ and $f_{j,k}$ are all continuous functions of $\alpha \in [0,1]$.

Finally, since

$$\begin{aligned} d_{j,k}(1 + e_{j,k} + f_{j,k}) &= d_{j,k} + d_{j,k}e_{j,k} + d_{j,k}f_{j,k} \\ &= E_{j,k} + B_{j,k} + D_{j,k} + F_{j,k} + G_{j,k} \\ &\qquad - b_{j,k}(1 + \alpha e_{j,k-1} + f_{j,k-1}) \\ &\qquad - c_{j,k}(1 + e_{j-1,k} + \alpha f_{j-1,k}) > 0 \end{aligned}$$

then

$$1 + e_{j,k} + f_{j,k} > 0.$$

The theorem is proved. ∎

Theorem 2.2. *If the matrix A satisfies (1.4) and*

$$\begin{aligned} &E_{j,k} + B_{j,k+1} + D_{j+1,k} + F_{j-1,k} + G_{j,k-1} \geq 0, \\ &\qquad (\text{ i.e., } A^T \text{ is diagonally dominant;}) \qquad 1 \leq j,k \leq n \\ &B_{j,k+1} \geq G_{j,k}, \qquad D_{j+1,k} \geq F_{j,k} \end{aligned} \tag{2.3}$$

and the matrix B is defined by (1.7) and (1.8), then there exists $\alpha^* \in (0,1)$ such that for $\alpha \in [0,\alpha^*]$ the matrix $A + B$ is positive real and satisfies

$$((A+B)u, u) > (1-\beta)(Au, u) \tag{2.4}$$

for some β satisfying $0 \leq \beta < 1$ and any real vector $u \neq 0$.

Note that (2.3) is weaker assumption than (1.5) but (1.5) implies (2.3).

Proof. For any real vector $u \neq 0$ we have

$$
\begin{aligned}
(Au, u) &= \sum_{j=1}^{n}\sum_{k=2}^{n} B_{j,k}u_{j,k-1}u_{j,k} + \sum_{j=2}^{n}\sum_{k=1}^{n} D_{j,k}u_{j-1,k}u_{j,k} \\
&\quad + \sum_{j=1}^{n-1}\sum_{k=1}^{n} F_{j,k}u_{j+1,k}u_{j,k} + \sum_{j=1}^{n}\sum_{k=1}^{n-1} G_{j,k}u_{j,k+1}u_{j,k} \\
&\quad + \sum_{j=1}^{n}\sum_{k=1}^{n} E_{j,k}u_{j,k}^2 \\
&= -\frac{1}{2}\sum_{j=1}^{n}\sum_{k=2}^{n} B_{j,k}\Big\{(u_{j,k-1} - u_{j,k})^2 - u_{j,k}^2 - u_{j,k-1}^2\Big\} \\
&\quad - \frac{1}{2}\sum_{j=2}^{n}\sum_{k=1}^{n} D_{j,k}\Big\{(u_{j-1,k} - u_{j,k})^2 - u_{j,k}^2 - u_{j-1,k}^2\Big\} \\
&\quad - \frac{1}{2}\sum_{j=1}^{n-1}\sum_{k=1}^{n} F_{j,k}\Big\{(u_{j+1,k} - u_{j,k})^2 - u_{j,k}^2 - u_{j+1,k}^2\Big\} \\
&\quad - \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n-1} G_{j,k}\Big\{(u_{j,k+1} - u_{j,k})^2 - u_{j,k}^2 - u_{j,k+1}^2\Big\} \\
&\quad + \sum_{j=1}^{n}\sum_{k=1}^{n} E_{j,k}u_{j,k}^2.
\end{aligned}
$$

Making some transformations of subscripts in the above expression we get

$$
\begin{aligned}
(Au, u) &= -\frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n-1}(B_{j,k+1} + G_{j,k})(u_{j,k+1} - u_{j,k})^2 \\
&\quad - \frac{1}{2}\sum_{j=1}^{n-1}\sum_{k=1}^{n}(D_{j+1,k} + F_{j,k})(u_{j+1,k} - u_{j,k})^2 \\
&\quad + \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n}(E_{j,k} + B_{j,k} + D_{j,k} + F_{j,k} + G_{j,k})u_{j,k}^2 \\
&\quad + \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n}(E_{j,k} + B_{j,k+1} + D_{j+1,k} + F_{j-1,k} + G_{j,k-1})u_{j,k}^2 > 0.
\end{aligned}
\tag{2.5}
$$

On the other hand, for any real vector $u \neq 0$ from (1.7) we have

$$
\begin{aligned}
(Bu,u) = & \sum_{j=1}^{n-1}\sum_{k=2}^{n} b_{j,k}e_{j,k-1}\Big\{u_{j+1,k-1} + \alpha u_{j,k} - \alpha u_{j+1,k} - \alpha u_{j,k-1}\Big\}u_{j,k} \\
& + \sum_{j=2}^{n}\sum_{k=1}^{n-1} c_{j,k}f_{j-1,k}\Big\{u_{j-1,k+1} + \alpha u_{j,k} - \alpha u_{j-1,k} - \alpha u_{j,k+1}\Big\}u_{j,k} \\
& = -\frac{1}{2}\sum_{j=1}^{n-1}\sum_{k=2}^{n} b_{j,k}e_{j,k-1}\Big[(u_{j,k} - u_{j+1,k-1})^2 - \alpha(u_{j+1,k} - u_{j,k})^2 \\
& \qquad - \alpha(u_{j,k-1} - u_{j,k})^2 - u_{j,k}^2 - u_{j+1,k-1}^2 + \alpha u_{j+1,k}^2 + \alpha u_{j,k-1}^2\Big] \\
& - \frac{1}{2}\sum_{j=2}^{n}\sum_{k=1}^{n-1} c_{j,k}f_{j-1,k}\Big[(u_{j,k} - u_{j-1,k+1})^2 - \alpha(u_{j-1,k} - u_{j,k})^2 \\
& \qquad - \alpha(u_{j,k+1} - u_{j,k})^2 - u_{j,k}^2 - u_{j-1,k+1}^2 + \alpha u_{j-1,k}^2 + \alpha u_{j,k+1}^2\Big].
\end{aligned}
$$

Varying the subscripts and using the convention $b_{j,1} = 0$ and $c_{1,k} = 0$ we get

$$
\begin{aligned}
(Bu,u) = & -\frac{1}{2}\sum_{j=1}^{n-1}\sum_{k=1}^{n-1}(b_{j,k+1}e_{j,k} + c_{j+1,k}f_{j,k})(u_{j,k+1} - u_{j+1,k})^2 \\
& + \frac{\alpha}{2}\sum_{j=1}^{n-1}\sum_{k=1}^{n-1}(b_{j,k}e_{j,k-1} + c_{j+1,k}f_{j,k})(u_{j+1,k} - u_{j,k})^2 \\
& \quad + \frac{\alpha}{2}\sum_{j=1}^{n-1} b_{j,n}e_{j,n-1}(u_{j+1,n} - u_{j,n})^2 \\
& + \frac{\alpha}{2}\sum_{j=1}^{n-1}\sum_{k=1}^{n-1}(b_{j,k+1}e_{j,k} + c_{j,k}f_{j-1,k})(u_{j,k+1} - u_{j,k})^2 \\
& \quad + \frac{\alpha}{2}\sum_{k=1}^{n-1} c_{n,k}f_{n-1,k}(u_{n,k+1} - u_{n,k})^2 \qquad (2.6) \\
& + \frac{1}{2}\sum_{j=1}^{n-1}\sum_{k=2}^{n} b_{j,k}e_{j,k-1}u_{j,k}^2 + \frac{1}{2}\sum_{j=2}^{n}\sum_{k=1}^{n-1} b_{j-1,k+1}e_{j-1,k}u_{j,k}^2 \\
& - \frac{\alpha}{2}\sum_{j=1}^{n-1}\sum_{k=1}^{n-1} b_{j,k+1}e_{j,k}u_{j,k}^2 - \frac{\alpha}{2}\sum_{j=2}^{n}\sum_{k=2}^{n} b_{j-1,k}e_{j-1,k-1}u_{j,k}^2 \\
& + \frac{1}{2}\sum_{j=2}^{n}\sum_{k=1}^{n-1} c_{j,k}f_{j-1,k}u_{j,k}^2 + \frac{1}{2}\sum_{j=1}^{n-1}\sum_{k=2}^{n} c_{j+1,k-1}f_{j,k-1}u_{j,k}^2 \\
& - \frac{\alpha}{2}\sum_{j=1}^{n-1}\sum_{k=1}^{n-1} c_{j+1,k}f_{j,k}u_{j,k}^2 - \frac{\alpha}{2}\sum_{j=2}^{n}\sum_{k=2}^{n} c_{j,k-1}f_{j-1,k-1}u_{j,k}^2.
\end{aligned}
$$

Dividing each of the last eight summations of the above expression into four parts, respectively, for example

$$\frac{1}{2}\sum_{j=1}^{n-1}\sum_{k=2}^{n} b_{j,k}e_{j,k-1}u_{j,k}^2 = \frac{1}{2}\Big\{\sum_{j=2}^{n-1}\sum_{k=2}^{n-1} b_{j,k}e_{j,k-1}u_{j,k}^2 + \sum_{k=2}^{n-1} b_{1,k}e_{1,k-1}u_{1,k}^2 + \sum_{j=2}^{n-1} b_{j,n}e_{j,n-1}u_{j,n}^2 + b_{1,n}e_{1,n-1}u_{1,n}^2\Big\}$$

and defining S_1, S_2, S_3 by

$$\begin{aligned}(S_1u,u) \equiv &-\frac{1}{2}\sum_{j=1}^{n-1}\sum_{k=1}^{n-1}(b_{j,k+1}e_{j,k}+c_{j+1,k}f_{j,k})(u_{j,k+1}-u_{j+1,k})^2 \\ &+\frac{\alpha}{2}\sum_{j=1}^{n-1}\sum_{k=1}^{n-1}(b_{j,k}e_{j,k-1}+c_{j+1,k}f_{j,k})(u_{j+1,k}-u_{j,k})^2 \\ &+\frac{\alpha}{2}\sum_{j=1}^{n-1}\sum_{k=1}^{n-1}(b_{j,k+1}e_{j,k}+c_{j,k}f_{j-1,k})(u_{j,k+1}-u_{j,k})^2 \\ &-\frac{\alpha}{2}(b_{1,2}e_{1,1}+c_{2,1}f_{1,1})u_{1,1}^2-\frac{\alpha}{2}(b_{n-1,n}e_{n-1,n-1}+c_{n,n-1}f_{n-1,n-1})u_{n,n}^2,\end{aligned} \tag{2.7}$$

$$\begin{aligned}(S_2u,u) \equiv &\frac{\alpha}{2}\sum_{j=1}^{n-1} b_{j,n}e_{j,n-1}(u_{j+1,n}-u_{j,n})^2+\frac{\alpha}{2}\sum_{k=1}^{n-1} c_{n,k}f_{n-1,k}(u_{n,k+1}-u_{n,k})^2 \\ &+\frac{1}{2}(b_{1,n}e_{1,n-1}+c_{2,n-1}f_{1,n-1})u_{1,n}^2+\frac{1}{2}(c_{n,1}f_{n-1,1}+b_{n-1,2}e_{n-1,1})u_{n,1}^2,\end{aligned} \tag{2.8}$$

and

$$\begin{aligned}(S_3u,u) \equiv &\frac{1}{2}\sum_{j=2}^{n-1}\sum_{k=2}^{n-1}\Big[(b_{j,k}e_{j,k-1}+b_{j-1,k+1}e_{j-1,k})-\alpha(b_{j-1,k}e_{j-1,k-1}+b_{j,k+1}e_{j,k})\Big]u_{j,k}^2 \\ &+\frac{1}{2}\sum_{j=2}^{n-1}\sum_{k=2}^{n-1}\Big[(c_{j,k}f_{j-1,k}+c_{j+1,k-1}f_{j,k-1})-\alpha(c_{j+1,k}f_{j,k}+c_{j,k-1}f_{j-1,k-1})\Big]u_{j,k}^2 \\ &+\frac{1}{2}\sum_{j=2}^{n-1}\Big[(b_{j-1,2}e_{j-1,1}-\alpha b_{j,2}e_{j,1})+(c_{j,1}f_{j-1,1}-\alpha c_{j+1,1}f_{j,1})\Big]u_{j,1}^2 \\ &+\frac{1}{2}\sum_{j=2}^{n-1}\Big[(b_{j,n}e_{j,n-1}-\alpha b_{j-1,n}e_{j-1,n-1})+(c_{j+1,n-1}f_{j,n-1}-\alpha c_{j,n-1}f_{j-1,n-1})\Big]u_{j,n}^2 \\ &+\frac{1}{2}\sum_{k=2}^{n-1}\Big[(b_{1,k}e_{1,k-1}-\alpha b_{1,k+1}e_{1,k})+(c_{2,k-1}f_{1,k-1}-\alpha c_{2,k}f_{1,k})\Big]u_{1,k}^2 \\ &+\frac{1}{2}\sum_{k=2}^{n-1}\Big[(b_{n-1,k+1}e_{n-1,k}-\alpha b_{n-1,k}e_{n-1,k-1})+(c_{n,k}f_{n-1,k}-\alpha c_{n,k-1}f_{n-1,k-1})\Big]u_{n,k}^2\end{aligned} \tag{2.9}$$

we have

$$(Bu,u) = (S_1u,u) + (S_2u,u) + (S_3u,u)\,. \tag{2.10}$$

Obviously, $(S_2u,u) \ge 0$. From Theorem 2.1 we know that for the grid point (j,k), $2 \le j,k \le n-1$,

$$b_{j,k}e_{j,k-1} + b_{j-1,k+1}e_{j-1,k} \ge \frac{B_{j,k}F_{j,k-1}}{d_{j,k-1}} + \frac{B_{j-1,k+1}F_{j-1,k}}{d_{j-1,k}} > 0.$$

This expression is a continuous function of $\alpha \in [0,1]$, so it has a positive minimum. On the other hand, $b_{j-1,k}e_{j-1,k-1} + b_{j,k+1}e_{j,k}$ is also a continuous function of $\alpha \in [0,1]$. Therefore, there exists some $\alpha_{j,k} \in (0,1)$ such that for $\alpha \in [0,\alpha_{j,k}]$ we have

$$(b_{j,k}e_{j,k-1} + b_{j-1,k+1}e_{j-1,k}) - \alpha\,(b_{j-1,k}e_{j-1,k-1} + b_{j,k+1}e_{j,k}) \ge 0.$$

Taking $\alpha_1 = \min_{2\le j,k\le n-1}\{\alpha_{j,k}\}$, if $\alpha \in [0,\alpha_1]$, the first summation of (S_3u,u) is not less than zero. Dealing with the other summations of (S_3u,u) in the same way, we can get $\alpha_2 \in (0,1)$ such that when $0 \le \alpha \le \alpha_2$, then

$$(S_3u,u) \ge 0. \tag{2.11}$$

In order to discuss (S_1u,u), we cite Lemma 3 in [2] here: let c and f be positive and let a, b, e be complex, then

$$\frac{cf}{c+f}|a-b|^2 \le f|a-e|^2 + c|b-e|^2.$$

So, for each grid point (j,k), $1 \le j,k \le n-1$, we have

$$\begin{aligned}
-\frac{1}{2}&b_{j,k+1}e_{j,k}(u_{j,k+1}-u_{j+1,k})^2\\
&= \frac{1}{2}\frac{b_{j,k+1}+d_{j,k}e_{j,k}}{d_{j,k}}\cdot\frac{d_{j,k}e_{j,k}\cdot b_{j,k+1}}{-b_{j,k+1}-d_{j,k}e_{j,k}}(u_{j,k+1}-u_{j+1,k})^2\\
&\ge -\frac{1}{2}\frac{b_{j,k+1}+d_{j,k}e_{j,k}}{d_{j,k}}\Big[b_{j,k+1}(u_{j,k+1}-u_{j,k})^2 + d_{j,k}e_{j,k}(u_{j+1,k}-u_{j,k})^2\Big]\\
&= -\frac{1}{2}\frac{b_{j,k+1}+d_{j,k}e_{j,k}}{d_{j,k}}\Big[(B_{j,k+1}-\alpha b_{j,k+1}e_{j,k})(u_{j,k+1}-u_{j,k})^2\\
&\qquad + (F_{j,k}-\alpha b_{j,k}e_{j,k-1})(u_{j+1,k}-u_{j,k})^2\Big].
\end{aligned} \tag{2.12}$$

But, by assumption (2.3) we have

$$\begin{aligned}
d_{j,k} + d_{j,k}e_{j,k} + b_{j,k+1} &= d_{j,k} + d_{j,k}e_{j,k} + B_{j,k+1} - \alpha b_{j,k+1}e_{j,k}\\
&\ge d_{j,k} + d_{j,k}e_{j,k} + G_{j,k} - \alpha b_{j,k+1}e_{j,k}\\
&= d_{j,k}(1 + e_{j,k} + f_{j,k}) + \alpha e_{j,k}f_{j-1,k} - \alpha b_{j,k+1}e_{j,k}.
\end{aligned}$$

Because $1 + e_{j,k} + f_{j,k}$ is a continuous function of $\alpha \in [0,1]$ and has a positive minimum, as in the above discussion, there exists $\alpha^o_{j,k} \in (0,1)$ such that when $0 \le \alpha \le \alpha^o_{j,k}$ we have

$$d_{j,k} + d_{j,k}e_{j,k} + b_{j,k+1} > 0$$

or

$$-\frac{b_{j,k+1} + d_{j,k}e_{j,k}}{d_{j,k}} = \beta_{j,k} < 1. \tag{2.13}$$

Set

$$\alpha_3 = \min_{1\le j,k\le n-1}\left\{\alpha^*_{j,k}\right\}$$

then, for $0 \le \alpha \le \alpha_3$ we get

$$0 \le \beta_1 = \max_{1\le j,k\le n-1}\left\{\beta_{j,k}\right\} < 1 \tag{2.14}$$

and

$$\begin{aligned}
-\frac{1}{2}\sum_{j=1}^{n-1}\sum_{k=1}^{n-1} & b_{j,k+1}e_{j,k}(u_{j+1,k}-u_{j,k+1})^2 \\
&\ge \frac{1}{2}\beta_1\sum_{j=1}^{n-1}\sum_{k=1}^{n-1}(B_{j,k+1}-\alpha b_{j,k+1}e_{j,k})(u_{j,k+1}-u_{j,k})^2 \\
&\quad + \frac{1}{2}\beta_1\sum_{j=1}^{n-1}\sum_{k=1}^{n-1}(F_{j,k}-\alpha b_{j,k}e_{j,k-1})(u_{j+1,k}-u_{j,k})^2.
\end{aligned} \tag{2.15}$$

In the same way we can prove that there exists $\alpha_4 \in (0,1)$ such that when $\alpha \in [0,\alpha_4]$ we have

$$0 \le \beta_2 < 1$$

and

$$\begin{aligned}
-\frac{1}{2}\sum_{j=1}^{n-1}\sum_{k=1}^{n-1} & c_{j+1,k}f_{j,k}(u_{j+1,k}-u_{j,k+1})^2 \\
&\ge \frac{1}{2}\beta_2\sum_{j=1}^{n-1}\sum_{k=1}^{n-1}(D_{j+1,k}-\alpha c_{j+1,k}f_{j,k})(u_{j+1,k}-u_{j,k})^2 \\
&\quad + \frac{1}{2}\beta_2\sum_{j=1}^{n-1}\sum_{k=1}^{n-1}(G_{j,k}-\alpha c_{j,k}f_{j-1,k})(u_{j,k+1}-u_{j,k})^2.
\end{aligned} \tag{2.16}$$

Let $\beta = \max(\beta_1,\beta_2)$. It is obvious that there exists $\alpha_5 \in (0,1)$ such that for $\alpha \in [0,\alpha_5]$ we have

$$\begin{aligned}
&\frac{1}{2}\beta(E_{1,1}+F_{1,1}+G_{1,1}) - \frac{1}{2}\alpha(b_{1,2}e_{1,1}+c_{2,1}f_{1,1}) \ge 0, \\
&\frac{1}{2}\beta(E_{n,n}+B_{n,n}+D_{n,n}) - \frac{1}{2}\alpha(b_{n-1,n}e_{n-1,n-1}+c_{n,n-1}f_{n-1,n-1}) \ge 0.
\end{aligned} \tag{2.17}$$

Finally, take $\alpha^* = \min\{\alpha_2, \alpha_3, \alpha_4, \alpha_5\}$. Then, if $0 \le \alpha \le \alpha^*$, we have

$$\begin{aligned}
(Au,u)+(Bu,u) \ge & -\frac{1}{2}(1-\beta)\sum_{j=1}^{n-1}\sum_{k=1}^{n-1}(B_{j,k+1}+G_{j,k})(u_{j,k+1}-u_{j,k})^2 \\
& -\frac{1}{2}(1-\beta)\sum_{j=1}^{n-1}\sum_{k=1}^{n-1}(D_{j+1,k}+F_{j,k})(u_{j+1,k}-u_{j,k})^2 \\
& -\frac{1}{2}\sum_{k=1}^{n-1}(B_{n,k+1}+G_{n,k})(u_{n,k+1}-u_{n,k})^2 \\
& -\frac{1}{2}\sum_{j=1}^{n-1}(D_{j+1,n}+F_{j,n})(u_{j+1,n}-u_{j,n})^2 \\
& +\frac{1}{2}\alpha(1-\beta)\sum_{j=1}^{n-1}\sum_{k=1}^{n-1}(b_{j,k}e_{j,k-1}+c_{j+1,k}f_{j,k})(u_{j+1,k}-u_{j,k})^2 \qquad (2.18)\\
& +\frac{1}{2}\alpha(1-\beta)\sum_{j=1}^{n-1}\sum_{k=1}^{n-1}(b_{j,k+1}e_{j,k}+c_{j,k}f_{j-1,k})(u_{j,k+1}-u_{j,k})^2 \\
& +\frac{1}{2}(1-\beta)\sum_{j=1}^{n-1}\sum_{k=1}^{n-1}(E_{j,k}+B_{j,k}+D_{j,k}+F_{j,k}+G_{j,k})u_{j,k}^2 \\
& +\frac{1}{2}\sum_{j=1}^{n-1}\sum_{k=1}^{n-1}(E_{j,k}+B_{j,k+1}+D_{j+1,k}+F_{j-1,k}+G_{j,k-1})u_{j,k}^2 \\
& +(S_2u,u)+(S_3u,u) \\
\ge & (1-\beta)(Au,u) > 0.
\end{aligned}$$

∎

3. Convergence conditions of SIP

Now we return to the case that the matrix A satisfies the conditions (1.4) and (1.5), so it is a symmetric positive definite matrix. First we quote Lemma 2.1 in [3] as one of our theorems.

Theorem 3.1. *The iterative scheme (1.9) is convergent if and only if*

$$Re\{\lambda_j(A^{-1}(A+B))\} > \frac{\tau}{2}, \qquad (3.1)$$

where $\lambda_j(W)$ denotes an eigenvalue of the matrix W, and $Re\,\mu$ denotes the real part of the complex μ.

Theorem 3.2. *If the matrix A satisfies (1.4), (1.5) and the matrix B is defined by (1.7) and (1.8), then there exist $\alpha^{**} \in (0,1)$ and $\beta^{**} \in [0,1)$ such that, for $0 \le \alpha \le \alpha^{**}$,*

$$Re\{\lambda_j(A^{-1}(A+B))\} \ge (1-\beta^{**}). \qquad (3.2)$$

Proof. As in the proof of Theorem 2.2, we can get $\alpha^{**} \in (0,1)$ and $\beta^{**} \in [0,1)$. If $0 \leq \alpha \leq \alpha^{**}$,

$$((A+B)u,u) \geq (1-\beta^{*})(Au,u), \tag{3.3}$$

where $u \neq 0$ is any real vector. Let λ_j be an eigenvalue of $A^{-1}(A+B)$ and $x+iy \neq 0$ be the associated eigenvector, i.e.

$$A^{-1}(A+B)(x+iy) = \lambda_j(x+iy).$$

So

$$((A+B)(x+iy), x+iy) = \lambda_j\,(A(x+iy), x+iy)\,.$$

Because A is symmetric positive definite,

$$((A+B)x,x) + ((A+B)y,y) = Re\lambda_j\,\{(Ax,x)+(Ay,y)\}\,.$$

From (3.3), we get

$$Re\lambda_j \geq (1-\beta^{**}).$$

∎

Based on Theorems 3.1 and 3.2 our main result is obtained.

Theorem 3.3. *If the matrix A satisfies (1.4), (1.5) and the matrix B is defined by (1.7) and (1.8), then exist $\alpha^{**} \in (0,1)$ and $\beta^{**} \in [0,1)$. When $0 \leq \alpha \leq \alpha^{**}$ and*

$$\tau < 2(1-\beta^{**}), \tag{3.4}$$

the iterative scheme (1.9) converges.

Acknowledgement. I would like to thank Martin H. Schultz for providing me an opportunity to visit Yale University and Longjun Shen and Howard C. Elman for their help.

References

[1] A. Bracha-Barak and P. E. Saylor,*A Symmetric Factorization Procedure for the Solution of Elliptic Boundary Value Problems*, SIAM J. Numer. Anal., 10(1973), pp. 190-206.
[2] T. Dupont, R. Kendall and H. H. Rachford Jr., *An approximate fatorization procedure for solving self-adjiont elliptic equations*, SIAM J. Numer. Anal., 5(1968), pp. 559-573.
[3] E. A. Lipitakis and D. J. Evans, *The Rate of Convergence of an Approximate Matix Factorization Semi-Direct Method*, Numer. Math., 36, 237-251(1981).
[4] H. L. Stone, *Iterative solution of implicit approximations of multidimensional partial differential equations*, SIAM J. Numer. Anal., 5(1968), pp. 530-558.
[5] D. M. Young, *Iterative solution of large linear systems*, New York-London: Academic Press, 1971.

DATE

ILMED

-8